

Ist einer von drei gleich jedem Dritten? Der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte

Lamp, Erich

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Lamp, E. (2001). Ist einer von drei gleich jedem Dritten? Der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 49, 49-68.

<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-199156>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der
Leibniz
Leibniz-Gemeinschaft

Ist einer von drei gleich jedem Dritten?

Der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte

von Erich Lamp¹

Zusammenfassung

Die Untersuchung knüpft an die in der Umfrageforschung und -praxis bekannte Tatsache an, dass logisch äquivalente Begriffsverwendungen in der Frageformulierung einen erheblichen Einfluss auf das Antwortverhalten der Befragten ausüben können. In einer Reihe von Feldexperimenten wird der Frage nachgegangen, inwieweit dies auch für numerische Äquivalente zutrifft. Anhand der systematischen Variation von Fragemodell, Frageformulierung, Frageinhalt und statistischer Größenordnung wird geprüft, ob die Verwendung verschiedener Mengenbezeichnungen zu verschiedenen Einschätzungen der mit den Mengenangaben verbundenen Sachverhalten führt.

Abstract

The present investigation is connected to the well known fact in public opinion research that the use of logically equivalent terms in question phrasing can lead to a considerable impact on interviewees' answering behavior. In a series of split ballot experiments the question of whether this also holds true in the case of numerical equivalents is pursued. By means of systematic variations of question patterns, wording, content and dimensions of quantities, this article tests whether the application of different notations of amount (numerical equivalents) could lead to different assessments of facts or circumstances that are connected with these nomenclatures.

¹ Dr. habil. **Erich Lamp** ist wissenschaftlicher Mitarbeiter im Institut für Publizistik, Johannes Gutenberg-Universität Mainz, Colonel-Kleinmann-Weg 2, D-55099 Mainz, erich.lamp@uni-mainz.de.

1. Einleitung

Von einem der Pioniere der amerikanischen Meinungsforschung, **Elmo Roper**, stammt der Vorschlag, systematisch Schlüsselbegriffe zu identifizieren, die im Fragebogen die subjektive Wahrnehmung und Bewertung des Frageinhalts durch die Befragten beeinflussen. Er illustriert seinen Vorschlag am Beispiel zweier Frageformulierungen, mit denen die Einstellung der Bevölkerung zur Rolle Amerikas bei der Befestigung des Weltfriedens untersucht wurde. Die Frage: "Do you think the United States should do everything in its power to promote world peace?", wurde von der überwältigenden Mehrheit der Befragten mit "Ja" beantwortet. Lautete die Frage hingegen: "Do you think the United States should become involved in a plan to promote world peace?", gingen die Ja-Stimmen in beträchtlichem Umfang zurück. Ganz offenbar, so **Roper**, handelt es sich bei dem Begriff "involved" um eines jener "danger words", von denen man kaum im Vorhinein wissen könne, wie sehr sie die Wahrnehmung und Bewertung eines Sachverhalts beeinflussen (**Roper** 1940).

In einem der frühen Standardwerke zur Umfrageforschung, "The Art of Asking Questions" von **Stanley Payne** aus dem Jahr 1951, findet sich bereits eine "Verbrechergalerie von Problemwörtern", die aus einer Liste von tausend im amerikanischen Sprachgebrauch häufig benutzten Alltagsbegriffen zusammengestellt wurde. Beispiele hierfür sind Begriffe wie "people", "government", "public", "any", "much" oder "few", die je nach Kontext oder Verständnis ganz verschiedene Bedeutungen annehmen können (**Payne** 1965). Am zuletzt genannten Begriff "few" lässt sich die Problematik sehr gut auch für den deutschen Sprachgebrauch veranschaulichen. In einer empirischen Untersuchung hat **Hörmann** nachgewiesen, dass mit dem Begriff "ein paar" ganz unterschiedliche Vorstellungen verbunden werden, je nachdem, ob die dadurch bezeichneten Objekte groß oder klein sind oder in welcher räumlichen Situation sie vorgestellt werden (**Hörmann** 1983).²

Zur Demonstration der Ergebnisverzerrungen, die durch die Wahl bestimmter Begriffe in der Frageformulierung entstehen können, wird in der Methodenliteratur häufig ein von **Rugg** berichtetes Fragebogenexperiment zitiert, in dem es um die Verwendung der Begriffe "erlauben" und "verbieten" geht (**Rugg** 1941; vgl. u.a. **Bradburn, Sudman** 1989; **Clark, Schober** 1992; **Friedrichs** 1973; **Noelle-Neumann, Petersen** 1996; **Reuband** 2001; **Schuman, Presser** 1981; **Sudman, Bradburn** 1982). Auf die Frage: "Do you think the United States should allow public speeches against democracy?", meinten 21 Prozent der Befragten, öffentliches Reden gegen die Demokratie sollte erlaubt sein, 62 Prozent vertraten

2 Versuchspersonen bezifferten beispielsweise die Anzahl von "ein paar Büroklammern" als deutlich mehr als die Anzahl von "ein paar Menschen". Diese wiederum wurden als deutlich mehr eingeschätzt, wenn sie statt "vor der Hütte" stehend als "vor dem Gebäude" stehend vorgestellt wurden. Auch änderte sich die geschätzte Anzahl von "ein paar" je nach der vorgestellten Größe des Blickfelds des Betrachters (Guckloch = weniger, Fenster = mehr) oder nach der Art des eingeführten Verbs: ein paar sich streitende Menschen sind weniger als lediglich ein paar Menschen.

die Ansicht, dies sollte nicht erlaubt sein. In der zweiten Version wurde der Begriff "erlauben" durch den Begriff "verbieten" ersetzt. Die Frage lautete: "Do you think the United States should forbid public speeches against democracy?" Auf diese Frageversion hin meinten 46 Prozent der Befragten, öffentliches Reden gegen die Demokratie sollte verboten sein, 39 Prozent plädierten dafür, dies nicht zu verbieten. Obwohl es sich, logisch gesehen, um identische Fragen handelt, sprechen sich einmal 62 Prozent für ein Verbot aus, das andere Mal 46 Prozent. **Rugg** zufolge liegt der Schlüssel für diesen Unterschied im Begriff "verbieten" (**Rugg** 1941), der offenbar eine viel schärfere Formulierung darstellt als das Gegenstück "nicht erlauben" (**Bradburn, Sudman** 1989). Nach **Noelle-Neumann** und **Petersen** (1996) vermeidet es ein Teil der Befragten, besonders scharf formulierten Positionen zuzustimmen, ebenso wie bei vielen eine Abneigung bestehe, auf eine Frage mit "Nein" zu antworten.

Ebenso können aber auch Fragen, deren Formulierungen keinerlei Schärfe besitzen, bei verschiedener Begriffsverwendung für identische Sachverhalte zu unterschiedlichen Ergebnissen führen, wie am Beispiel der Begriffe "aufwachen" und "durchschlafen" deutlich wird.³ Befragte antworten auf die Frage "Wachen Sie nachts schon mal auf?" deutlich häufiger mit "Wache manchmal auf" als auf die – alternativ formulierte – Frage "Schlafen Sie nachts in der Regel durch?". Sie geben deutlich seltener an, "nie" aufzuwachen als "immer" durchzuschlafen (Tabelle 1). Offensichtlich entscheiden sich die Befragten hier häufiger für Antwortalternativen, deren Formulierung derjenigen der Testfrage ähnelt als für Alternativen, die sich sprachlich davon unterscheiden – obwohl die semantische Bedeutung nach allgemeinem Sprachgefühl dieselbe ist.

3 Das Beispiel stammt, ebenso wie die unten berichteten Ergebnisse, aus gegabelten face-to-face-Befragungen der Mainzer Bevölkerung, die im Methodenabschnitt näher erläutert werden.

Ersatzseite für ZA-Information 49, Seite 51: (letzter Absatz dort fehlend)

die Ansicht, dies sollte nicht erlaubt sein. In der zweiten Version wurde der Begriff "erlauben" durch den Begriff "verbieten" ersetzt. Die Frage lautete: "Do you think the United States should forbid public speeches against democracy?" Auf diese Frageversion hin meinten 46 Prozent der Befragten, öffentliches Reden gegen die Demokratie sollte verboten sein, 39 Prozent plädierten dafür, dies nicht zu verbieten. Obwohl es sich, logisch gesehen, um identische Fragen handelt, sprechen sich einmal 62 Prozent für ein Verbot aus, das andere Mal 46 Prozent. **Rugg** zufolge liegt der Schlüssel für diesen Unterschied im Begriff "verbieten" (**Rugg** 1941), der offenbar eine viel schärfere Formulierung darstellt als das Gegenstück "nicht erlauben" (**Bradburn, Sudman** 1989). Nach **Noelle-Neumann** und **Petersen** (1996) vermeidet es ein Teil der Befragten, besonders scharf formulierten Positionen zuzustimmen, ebenso wie bei vielen eine Abneigung bestehe, auf eine Frage mit "Nein" zu antworten.

Ebenso können aber auch Fragen, deren Formulierungen keinerlei Schärfe besitzen, bei verschiedener Begriffsverwendung für identische Sachverhalte zu unterschiedlichen Ergebnissen führen, wie am Beispiel der Begriffe "aufwachen" und "durchschlafen" deutlich wird.³ Befragte antworten auf die Frage "Wachen Sie nachts schon mal auf?" deutlich häufiger mit "Wache manchmal auf" als auf die – alternativ formulierte – Frage "Schlafen Sie nachts in der Regel durch?". Sie geben deutlich seltener an, "nie" aufzuwachen als "immer" durchzuschlafen (Tabelle 1). Offensichtlich entscheiden sich die Befragten hier häufiger für Antwortalternativen, deren Formulierung derjenigen der Testfrage ähnelt als für Alternativen, die sich sprachlich davon unterscheiden – obwohl die semantische Bedeutung nach allgemeinem Sprachgefühl dieselbe ist.

Über die Problematik von "danger words" hinaus lenken die angeführten Beispiele die Aufmerksamkeit auf die Bedeutung logisch äquivalenter Ausdrücke und Formulierungen, die unabhängig vom Frageinhalt zu Ergebnisverzerrungen führen können. Eine von **Kahnemann** und **Tversky** im Rahmen der Risikokommunikation durchgeführte Untersuchung liefert hierfür ein bemerkenswertes Beispiel. Die Befragten sollten in einem fiktiven Szenario zwischen zwei wissenschaftlichen Programmen entscheiden, mit denen eine ungewöhnliche asiatische Krankheit, die voraussichtlich 600 Menschen das Leben kosten würde, bekämpft werden sollte. Beide Programme führten im Prinzip zum gleichen Ergebnis; allerdings wurde dieses Ergebnis in einem Fall als sicher ("200 Menschen werden gerettet"), im anderen Fall als risikobehaftet dargestellt ("mit ein-Drittel-Wahrscheinlichkeit werden 600 Menschen gerettet, mit zwei-Drittel-Wahrscheinlichkeit wird niemand gerettet"). Die klare Mehrheit der Befragten (72 Prozent) bevorzugte das Programm, wonach die Rettung von 200 Menschen als gesichert galt, nur jeder Vierte (28 Prozent) war bereit, ein Risiko einzugehen, um alle 600 zu retten.

3 Das Beispiel stammt, ebenso wie die unten berichteten Ergebnisse, aus gegabelten face-to-face-Befragungen der Mainzer Bevölkerung, die im Methodenabschnitt näher erläutert werden.

Tabelle 1: Der Einfluss logischer Äquivalente auf das Antwortverhalten

Fragen: *"Wachen Sie nachts schon mal auf? Würden Sie sagen, das kommt häufiger vor oder nur manchmal, oder kommt das nie vor?"*

"Schlafen Sie nachts in der Regel durch? Würden Sie sagen, Sie schlafen immer durch, oder wachen Sie manchmal oder auch häufiger in der Nacht auf?"

	"Wachen Sie nachts schon mal auf?" %	"Schlafen Sie nachts in der Regel durch?" %
Wache nachts häufiger auf	23	24
Wache nachts manchmal auf	51	37
Wache nachts nie auf / Schlafe nachts immer durch	26	39
	<hr/> 100	<hr/> 100
n =	90	88

IfP-MP WS 98/99

Einer zweiten Gruppe von Befragten wurde der identische Sachverhalt von der anderen Seite her vorgestellt, nicht der Gewinn ("Leben retten") stand im Vordergrund, sondern der Verlust ("Leben verlieren"). Der Verlust des Lebens wurde bei Anwendung des einen Programms als unvermeidbar und sicher gegeben dargestellt ("400 Menschen werden sterben"), bei Anwendung des anderen Programms wiederum als mit einem Risiko behaftet ("mit ein-Drittel-Wahrscheinlichkeit wird niemand sterben, mit zwei-Drittel-Wahrscheinlichkeit werden alle ihr Leben verlieren"). In dieser Version der Fragestellung kehrten sich die Ergebnisse hinsichtlich der Risikobereitschaft der Befragten geradezu um. Die klare Mehrheit (78 Prozent) nahm nun lieber das Risiko in Kauf, um alle 600 Menschen zu retten, und nur jeder Fünfte (22 Prozent) entschied sich für das Programm, das den sicheren Tod von 400 Personen bedeutete (**Kahnemann, Tversky** 1984).

Logische Äquivalente sind ein beliebtes Stilmittel der Rhetorik, wenn es darum geht, Sachverhalte positiv oder negativ zu konnotieren. Beispiele dieser Art – nach dem Muster: das Glas ist halb voll / halb leer – sind jedem geläufig. In der Umfrageforschung ist bei Verwendung solcher Formulierungen mithin höchste Vorsicht geboten, nicht zuletzt wegen der (möglichen) Unterstellung manipulativer Absichten. So unabdingbar ein ausgeprägtes Sprachgefühl hierbei sein mag, so wenig reicht es hin, die Vielfalt der auf die Sprachverwendung zurückgehenden Einflüsse auf die subjektive Wahrnehmung und Bewertung des Frageinhalts zu ermessen. In vielen Fällen kann eine Ergebnisverzerrung durch Verwendung äquivalenter Ausdrücke und Formulierungen zwar vermutet werden. In mindestens

ebenso vielen Fällen, wenn nicht in den meisten, entzieht sich ein solcher Einfluss aber selbst sogar der Vorstellung. Dies ist insbesondere bei numerischen Äquivalenten der Fall.

Der Gedanke, dass die sprachliche Variation bei der Beschreibung von Mengenverteilungen nicht nur Abwechslung schafft und eine bessere Lesbarkeit des Textes bedeutet, sondern je nach Verwendung der numerischen Äquivalente auch ganz unterschiedliche Vorstellungen über die Größenordnung der beschriebenen Sachverhalte hervorruft, dürfte selbst geübten Autoren fernliegen. Wie selbstverständlich werden bei der Darstellung statistischer Sachverhalte, Ergebnissen von Meinungsumfragen etwa, Prozentwerte mit solchen Mengenbezeichnungen umschrieben: jeder Zweite, ein Drittel, einer von vier, jeder Fünfte usw., und kaum jemand setzt sie beim Lesen oder Hören wohl in nackte Prozentzahlen um. Welche Mengenmuster werden dabei im Vorstellungsraum abgerufen? Was trägt sich vor dem geistigen Auge zu, wenn von "jedem Dritten" die Rede ist, von "33 Prozent", von "einem Drittel" oder "einem von dreien"? Macht es für die Wahrnehmung und Bewertung eines Sachverhalts einen Unterschied, ob ein Bevölkerungsanteil unpersönlich abstrakt oder anschaulich konkret beschrieben wird? Dieser Problemstellung soll im Folgenden nachgegangen werden.

2. Methodisches Vorgehen

Die berichteten Ergebnisse stammen sämtlich aus gegabelten face-to-face-Befragungen der Mainzer Bevölkerung, die nach ihrer Logik dem kontrollierten Experiment folgen. Alle Umstände im Ablauf der Untersuchung sind gleich, bis auf einen, den experimentellen Faktor, der im Fragebogen variiert wird. Ziel einer solchen Anlage ist es, für Unterschiede, die nach Einwirkung des experimentellen Faktors gemessen werden, alle Erklärungen bis auf eine auszuschließen. Bei den regelmäßig jedes Semester durchgeführten Umfragen stehen insbesondere methodische Aspekte des Fragebogens und der Frageformulierung im Vordergrund des Interesses, um das Entdeckungspotential der Umfrageforschung durch Anwendung der Technik der gegabelten Befragung systematisch zu nutzen.

Die Fragebögen enthalten in der Regel Fragen zu ganz unterschiedlichen Themen und Sachverhalten und decken in ihrer Konstruktion die ganze Bandbreite der verschiedenartigen Fragemodelle und Hilfsmittel einschließlich Listen, Kartenspielen und Bildblattvorlagen ab. Die besondere Aufmerksamkeit gilt hierbei vor allem drei Gruppen von Einflussgrößen, die sich erstens auf die Anordnung und Abfolge der Fragen im Fragebogen beziehen, zweitens auf den Einsatz verschiedener Fragemodelle und -instrumente sowie drittens auf die Art der Frageformulierung. Die Interviews wurden von Studenten der Publizistikwissenschaft der Universität Mainz durchgeführt. Die Stichproben wurden nach dem Quotenverfahren gebildet, wobei die Merkmale Geschlecht und Alter (in der Verteilung 16-29, 30-44, 45-59, 60+ Jahre) vorgegeben waren.

Die Antworten der befragten Mainzer Bürger sind im statistischen Sinne nicht repräsentativ für die Gesamtbevölkerung der Bundesrepublik. Dies ist für die vorliegende Untersuchung unerheblich, da es nicht um die Meinungen aller Deutschen geht sondern darum, die Antworten von Befragten, denen unterschiedliche Fragevariationen vorgegeben wurden, systematisch miteinander zu vergleichen. Es gibt keinen Grund anzunehmen, dass Unterschiede, die sich in den Antworten der Mainzer Bürger zeigen, nicht auch für die Gesamtbevölkerung gelten.

3. Ergebnisse

Zur Überprüfung der Frage, ob die Wahl bestimmter Mengenbezeichnungen die Wahrnehmung und Bewertung eines Sachverhalts, der eine Quantifizierung enthält, beeinflusst, wurde folgende Testfrage entwickelt: *"Wenn Sie hören, ein Drittel der Arbeitnehmer meldet sich schon mal krank, ohne wirklich krank zu sein. Würden Sie sagen, das stellt eine Bedrohung für unsere Wirtschaft dar, oder sehen Sie darin keine ernste Gefahr?"* Im split ballot-Verfahren wurde diese Testfrage vierfach variiert: Der ansonsten gleiche Fragetext unterschied sich lediglich in der Mengenbezeichnung. Der Anteil der Arbeitnehmer wurde dabei vorgestellt als "ein Drittel", "jeder dritte", "etwa 33 Prozent" und "einer von drei". Die Einschätzung der Bedrohung der Wirtschaft erfolgte anhand einer vierstufigen Skala, die den Befragten zur optischen Unterstützung als Bildblatt mit vier unterschiedlich großen Quadraten vorgelegt wurde. Zur Erläuterung hieß es dazu im Fragetext weiter: *"Wenn Sie es mir bitte nach diesem Bildblatt hier sagen, es geht so: Der obere große Kasten würde bedeuten, Sie empfinden das als äußerst bedrohlich, der unterste kleine Kasten würde bedeuten, Sie sehen darin keine große Gefahr."*

Die Antworten der Befragten unterscheiden sich deutlich, je nachdem welche Mengenbezeichnung ihnen vorgelegt wurde. Fasst man die beiden oberen Skalenpunkte (außerordentlich große Gefahr / große Gefahr) und die beiden unteren Skalenpunkte (eher geringe Gefahr / geringe Gefahr) zusammen, unterscheiden sich die Ergebnisse um bis zu 11 Prozentpunkte. Die Befragten schätzten die Bedrohung für die Wirtschaft bei der Mengenangabe "jeder dritte Arbeitnehmer" als am größten ein, gefolgt von der Angabe "ein Drittel der Arbeitnehmer" und "einer von drei Arbeitnehmern". Den Eindruck der geringsten Gefährdung machte die Mengenangabe "etwa 33 Prozent der Arbeitnehmer", bei der sich die Mehrheitsverhältnisse sogar umkehren (Tabelle 2).

Tabelle 2: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:

Variation der Mengenbezeichnung "ein Drittel"

(Viererskala mit optischer Bildblatt-Unterstützung)

Frage: "Wenn Sie hören, ein Drittel der Arbeitnehmer meldet sich schon mal krank, ohne wirklich krank zu sein. Würden Sie sagen, das stellt eine Bedrohung für unsere Wirtschaft dar, oder sehen Sie darin keine ernste Gefahr? Wenn Sie es mir bitte nach diesem Bildblatt hier sagen, es geht so: Der obere große Kasten würde bedeuten, Sie empfinden das als äußerst bedrohlich, der unterste kleine Kasten würde bedeuten, Sie sehen darin keine große Gefahr?" (Vorlage eines Bildblatts)

In Parallelbefragungen hieß es: "jeder dritte Arbeitnehmer" / "etwa 33 Prozent der Arbeitnehmer" / "einer von drei Arbeitnehmern".

	Anteil der Arbeitnehmer vorgestellt als -			
	jeder Dritte %	ein Drittel %	einer von drei %	etwa 33 Prozent %
Bedrohung (Stufen 4 und 3)	56	54	50	45
Keine große Gefahr (Stufen 2 und 1)	44	46	47	49
Unentschieden	---	---	3	6
	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
n =	59	61	60	51

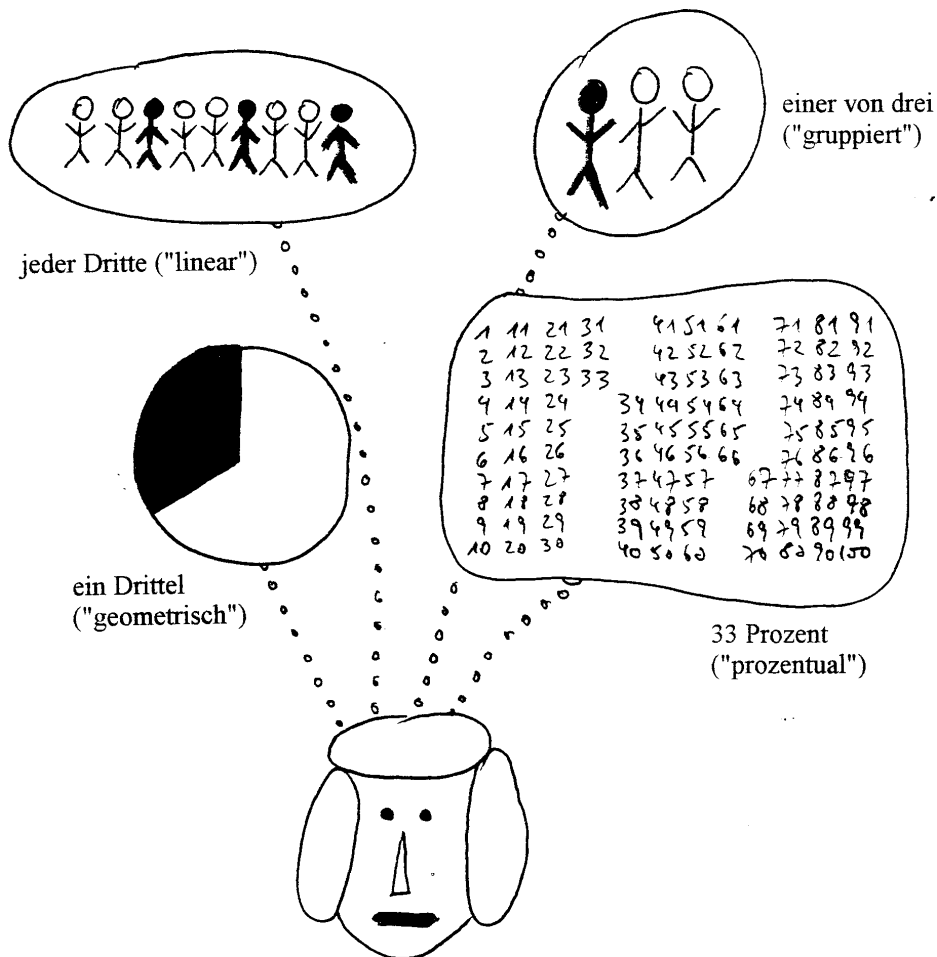
IfP-MP WS 95/96

Handelt es sich bei diesem Ergebnis um einen Zufallsbefund oder gibt es systematische Unterschiede in der Vorstellung von Mengen, die mit verschiedenen Bezeichnungen umschrieben werden? Neuere Untersuchungen aus dem Forschungsbereich der sogenannten "mental arithmetics" (**Besner, Coltheart** 1979; **Campbell, Clark** 1988, 1992; **Clark, Campbell** 1991; **Dehaene** 1999; **Dehaene, Cohen** 1991; **Gonzales, Kolers** 1982, 1987; **Hinrichs, Yurko, Hu** 1981; **Hinrichs, Berie, Mosell** 1982; **Vorberg, Blankenberger** 1993) haben gezeigt, dass Zahlen nicht kognitiv abstrakt sind, d.h. eine einheitliche mentale Repräsentation besitzen, sondern je nach Darstellungsform in verschiedene mentale Repräsentationen überführt werden können. So stellten **Besner** und **Coltheart** (1979) fest, dass Größenvergleiche von Zahlen in arabischer Darstellung nicht denselben Gesetzmäßigkeiten folgen wie Größenvergleiche von Zahlwörtern. Entsprechende Befunde werden für das Addieren von Zahlen, die in arabischen oder in römischen Ziffern dargeboten wurden, berichtet (**Gonzales, Kolers** 1982, 1987). Experimente mit Zahlen von 1 bis 6, die als arabische Ziffer, als Wort, als Augenmuster auf einem Würfel oder als zufälliges Punktmuster

dargeboten wurden, belegen ebenfalls eine mehrfache mentale Repräsentation von Zahlen (*Vorberg, Blankenberger* 1993).

Informelle Ermittlungen des Verfassers, die im Zusammenhang mit der hier behandelten Problemstellung bei zahlreichen Personen durchgeführt wurden, weisen in eine ähnliche Richtung. Nach Auskunft der Befragten stellen sich vor dem geistigen Auge je nach verwendeter Mengenbezeichnung durchaus verschiedene "Bilder" ein. Die mit den Mengenbezeichnungen spontan verknüpften Vorstellungen einer Vielzahl von Befragten lassen sich wie folgt zusammenfassen: Die Mengenbezeichnung "jeder Dritte" wird als eine Reihung von Personen oder Objekten gesehen, in der jedes dritte Element hervorgehoben ist. Demgegenüber erscheint die Mengenbezeichnung "einer von drei" als eine Gruppe von Personen oder Objekten, in der ein Element hervorgehoben ist. Mit der Mengenbezeichnung "ein Drittel" wiederum wird eher eine geometrische Figur verbunden, aus der ein Teilstück hervorsticht. Die Mengenbezeichnung "33 Prozent" verbleibt dagegen weitgehend im Bereich der Abstraktion und wird als zahlenmäßiger Anteil gesehen. Diese im Vorstellungsraum aktivierten Mengenmuster treffen auch auf die numerischen Äquivalente anderer Größenordnungen zu. Zum Zweck der sprachlichen Abgrenzung der Mengenbezeichnungen über alle Größenordnungen hinweg können sie – im Rückgriff auf die mentalen Repräsentationen – auch als "lineare" (jeder Dritte, jeder Vierte, jeder Fünfte usw.), "gruppierte" (einer von drei, einer von vier, einer von fünf usw.), "geometrische" (ein Drittel, ein Viertel, ein Fünftel usw.) und "prozentuale" (33 Prozent, 25 Prozent, 20 Prozent usw.) Mengenangaben bezeichnet werden (Schaubild).

Schaubild: Welche Mengenmuster werden bei Verwendung numerischer Äquivalente für identische Größenordnungen im Vorstellungsraum abgerufen?



Anmerkung: Die dargestellten Mengenmuster gehen auf informelle Ermittlungen des Verfassers zurück, die bei zahlreichen Personen durchgeführt wurden. Dargestellt sind die zumeist genannten "Bilder", die sich vor dem geistigen Auge der Befragten bei den o.g. Mengenbezeichnungen einstellten. Die im Vorstellungsraum aktivierten Mengenmuster treffen auch für die numerischen Äquivalente anderer Größenordnungen, z.B. ein Viertel, einer von vier usw. oder ein Fünftel, einer von fünf usw., zu.

Wenngleich den informellen Ermittlungen der im Vorstellungsraum abgerufenen Mengenmuster der numerischen Äquivalente keine Beweiskraft im Sinne der klassischen Testtheorie zukommt (vgl. dazu **Lienert, Raatz** 1994), helfen sie doch, den bis hierhin erst vage erkennbaren vermuteten Sachverhalt weiter auszuleuchten. Könnte es sein, dass die verschiedenen Mengenbezeichnungen aufgrund unterschiedlicher mentaler Repräsentationen voneinander abweichende semantische (arithmetische) und andere Informationen enthalten, die bei den Befragten unterschiedliche kognitive Prozesse auslösen (vgl. **Vorberg**,

Blankenberger 1993)? In diesem Fall sollten sich die Antworten der Befragten auch bei einem anderen Fragemodell je nach vorgelegter Mengenbezeichnung unterscheiden.

Die modifizierte Testfrage entspricht im Wortlaut exakt dem ersten Teil der oben vorgestellten Testfrage für die Größenordnung "ein Drittel". In Abänderung des Fragemodells erfolgte die Einschätzung der Bedrohung für die Wirtschaft (durch Krankheit simulierende Arbeitnehmer) nicht anhand einer vierstufigen Skala, die den Befragten zur optischen Unterstützung als Bildblatt vorgelegt wurde, sondern – entsprechend der Fragestellung – in Form dichotomisierter Antwortkategorien. Unabhängig von den vier Mengenbezeichnungen sehen die Befragten den Sachverhalt insgesamt häufiger als eine Bedrohung für die Wirtschaft an als dass sie ihn für gefahrlos halten. Bei den Mengenangaben "ein Drittel" und "einer von drei" fallen die Einschätzungen einer Bedrohung für die Wirtschaft jedoch deutlich höher aus als bei den Mengenbezeichnungen "33 Prozent" und "jeder dritte". Je nachdem, welche Mengenbezeichnung den Befragten vorgelegt wurde, unterscheiden sich die Einschätzungen um bis zu 12 Prozentpunkte (Tabelle 3). Die Tatsache, dass die Ergebnisverteilung bei Verwendung dichotomisierter Antwortkategorien in den Versuchsgruppen stärker polarisiert als bei Verwendung skalierteter Antwortkategorien (vgl. Tabelle 2), ist hierbei unerheblich. Dies kann neben dem Einfluss des anderen Fragemodells⁴ ebenso auf eine zum zweiten Messzeitpunkt tatsächlich veränderte Einstellung der Bevölkerung dem erfragten Sachverhalt gegenüber zurückzuführen sein.⁵

4 Wie **Tennstädt** und **Hansen** (1982) zeigen konnten, hat eine geringfügige Änderung des Fragemodells eine zehnmal größere Veränderung der Ergebnisse zur Folge als ein Wechsel der Stichproben-Technik.

5 Keineswegs unerheblich ist dieser Sachverhalt unter einem anderen methodischen Gesichtspunkt. Sollen die Meinungen und Einstellungen der Bevölkerung zu einer Frage als Trendanalyse im Zeitverlauf untersucht werden, ist die Beibehaltung des Fragemodells wie auch des Frage-Umfelds im Fragebogen unabdingbar, da sonst gleiche oder andere Ergebnisse zu verschiedenen Messzeitpunkten nichts mehr über Stabilität oder Wandel der Bevölkerungsmeinung aussagen.

Tabelle 3: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:

Variation der Mengenbezeichnung "ein Drittel"

(Dichotomisierte Antwortkategorien ohne Bildblatt-Unterstützung)

Frage: "Wenn Sie hören, ein Drittel der Arbeitnehmer meldet sich schon mal krank, ohne wirklich krank zu sein. Würden Sie sagen, das stellt eine Bedrohung für unsere Wirtschaft dar, oder sehen Sie darin keine ernste Gefahr?"

In Parallelbefragungen hieß es: "jeder dritte Arbeitnehmer" / "etwa 33 Prozent der Arbeitnehmer" / "einer von drei Arbeitnehmern".

	Anteil der Arbeitnehmer vorgestellt als -			
	ein Drittel ("geometrisch")	einer von drei ("gruppiert")	etwa 33 Prozent ("prozentual")	jeder dritte ("linear")
	%	%	%	%
Bedrohung	63	60	53	52
Keine große Gefahr	34	28	39	40
Unentschieden	3	12	8	8
	$\overline{100}$	$\overline{100}$	$\overline{100}$	$\overline{100}$
n =	71	50	57	60

IfP-MP WS 99/00

Das Ergebnis macht deutlich, dass die verschiedenen Mengenbezeichnungen offenbar unabhängig vom Fragemodell differenzieren. Dies spricht für die Vermutung, dass die numerischen Äquivalente für die Größenordnung "ein Drittel" unterschiedliche mentale Repräsentationen ("geometrisch", "gruppiert", "prozentual", "linear") besitzen, die bei den Befragten – aufgrund der verschiedenen hiermit verknüpften Informationen – unterschiedliche kognitive Prozesse (Einschätzung der Bedrohung) auslösen. Über Art und Bedeutung der mit den mentalen Repräsentationen verbundenen semantischen (mengenmäßigen) und anderen Informationen (z.B. benachbarte Größenordnungen) kann an dieser Stelle noch keine Aussage getroffen werden.⁶ Vielmehr soll zunächst dem Effekt "an sich" nachgegangen und geprüft werden, ob er auch für andere Größenordnungen und bei anderen Fragestellungen Geltung besitzt.

Zu diesem Zweck wird das zuletzt verwendete (vereinfachte) Fragemodell beibehalten und – für den Test der Mengenbezeichnungen einer anderen Größenordnung – auch der im

6 Dies betrifft die Frage, warum die getesteten Mengenbezeichnungen unabhängig vom Fragemodell differenzieren, ihre Differenzierung aber nicht einheitlich in der Weise erfolgt, dass z.B. mit der "linearen" Mengenbezeichnung hier wie da die Einschätzung der größten Bedrohung für die Wirtschaft verbunden ist.

Fragetext bisher angesprochene Sachverhalt. Lediglich der Anteil der in Frage stehenden Arbeitnehmer wurde variiert und in den vier Versuchsgruppen vorgestellt als "jeder zweite", "die Hälfte", "50 Prozent" sowie "einer von zwei". Die Ergebnisse untermauern die bisherigen Befunde. Die Wahrnehmung und Bewertung des identischen Sachverhalts durch die Befragten fällt je nach vorgelegter Mengenbeschreibung ganz verschieden aus. Bei den Mengenangaben "die Hälfte" und "jeder zweite" schätzen die Befragten die Bedrohung für die Wirtschaft deutlich größer ein als bei den Mengenbezeichnungen "50 Prozent" und "einer von zwei". Insgesamt unterscheiden sich die Einschätzungen der Versuchspersonen bei der Variation dieser Mengenbezeichnungen um bis zu 21 Prozentpunkte (Tabelle 4).

Tabelle 4: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:
Variation der Mengenbezeichnung "die Hälfte"

Frage: *"Wenn Sie hören, die Hälfte der Arbeitnehmer meldet sich schon mal krank, ohne wirklich krank zu sein. Würden Sie sagen, das stellt eine Bedrohung für unsere Wirtschaft dar, oder sehen Sie darin keine ernste Gefahr?"*

In Parallelbefragungen hieß es: *"jeder zweite Arbeitnehmer" / "50 Prozent der Arbeitnehmer" / "einer von zwei Arbeitnehmern"*.

	Anteil der Arbeitnehmer vorgestellt als -			
	die Hälfte ("geometrisch") %	jeder zweite ("linear") %	50 Prozent ("prozentual") %	einer von zwei ("gruppiert") %
Bedrohung	58	53	40	37
Keine große Gefahr	35	45	42	52
Unentschieden	7	2	18	11
	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
n =	40	40	38	35

IfP-MP SS 2000

Unabhängig vom Fragemodell scheinen die verschiedenen Mengenbezeichnungen nach diesem Ergebnis offenbar auch unabhängig von der vorgestellten Größenordnung zu differenzieren. Handelt es sich bei den numerischen Äquivalenten mithin um eine konstante Einflussgröße, die aufgrund der verschiedenen mentalen Repräsentationen die Urteile und Antworten der Befragten auch unabhängig vom Frageinhalt beeinflusst? In diesem Fall sollten die beobachteten Ergebnisverzerrungen auch bei anderen Frageformulierungen für andere Sachverhalte bei anderen als den bisher getesteten Größenordnungen auftreten. Zur Überprüfung dieser Annahmen wurde eine neue Testfrage entwickelt. Die Frage lautete:

"Kürzlich war zu lesen, jeder vierte Deutsche hätte zu wenig Selbstvertrauen. Finden Sie, das ist viel, oder ist das nicht so viel?" In den Parallelbefragungen wurde der Anteil der Deutschen entsprechend als "25 Prozent", "einer von vier" sowie "ein Viertel" vorgestellt. Die Testfrage wurde überdies unverändert in vier Befragungen mit unterschiedlichem Frageumfeld aufgenommen, um einen möglichen Kontexteffekt auszuschließen. Erneut zeigt sich auch bei veränderter Fragestellung und anderer Größenordnung der gleiche Effekt. In allen vier Befragungen unterscheiden sich die Antworten der Befragten deutlich je nach vorgelegter Mengenbezeichnung. Die zusammengefassten Daten zeigen, dass insbesondere die Mengenangabe "jeder vierte" bei den Befragten zu dem Eindruck führt, dies sei viel, wohingegen bei den Mengenangaben "einer von vier" und "25 Prozent" der Eindruck überwiegt, dies sei nicht so viel. Insgesamt fallen die Einschätzungen der Befragten bei der Variation dieser Mengenbezeichnung um bis zu 15 Prozentpunkte auseinander (Tabelle 5).

Tabelle 5: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:
Variation der Mengenbezeichnung "ein Viertel"

Frage: "Kürzlich war zu lesen, jeder vierte Deutsche hätte zu wenig Selbstvertrauen. Finden Sie, das ist viel, oder ist das nicht so viel?"

In Parallelbefragungen hieß es: "ein Viertel der Deutschen" / "25 Prozent der Deutschen" / "einer von vier Deutschen".

	Anteil der Deutschen vorgestellt als -			
	jeder vierte ("linear") %	ein Viertel ("geometrisch") %	einer von vier ("gruppiert") %	25 Prozent ("prozentual") %
Finde ich viel	49	45	38	34
Nicht so viel	40	47	53	54
Unentschieden	11	8	9	12
	<u>100</u>	<u>100</u>	<u>100</u>	<u>100</u>
n =	232	224	261	238

IfP-MP, Umfragen SS 1999, WS 99/00, SS 2000, WS 00/01 (zusammengefasste Daten)

Alle bisher vorgestellten Ergebnisse weisen in dieselbe Richtung: Numerische Äquivalente führen je nach Art der Mengenbezeichnung zu einer unterschiedlichen Wahrnehmung und Bewertung identischer Sachverhalte durch die Befragten, und zwar unabhängig von Fragemodell und -formulierung, Frageinhalt und statistischer Größenordnung. Lassen sich die Befunde an einer weiteren Größenordnung und einem wiederum anderen Sachverhalt erhärten? Die hierzu entwickelte Testfrage lautete: "Kürzlich war zu lesen, dass bei uns in

Deutschland jede fünfte Frau raucht. Finden Sie, das ist viel, oder ist das nicht so viel?"

In den Parallelumfragen wurde der Anteil der Frauen vorgestellt als "ein Fünftel", "eine von fünf" sowie "20 Prozent". Wie erwartet unterscheiden sich die Urteile und Antworten der Befragten auch in dieser Frageversion trotz des identischen Frageinhalts je nach vorgestellter Mengenbezeichnung. Auf die Mengenangaben "jede fünfte" und "20 Prozent" antworten die Befragten deutlich häufiger mit "Finde ich viel" als mit "Nicht so viel", wohingegen sich die Einschätzungen bei den Mengenangaben "ein Fünftel" und "eine von fünf" die Waage halten. Die maximale Differenz der Angaben in den einzelnen Versuchsgruppen beträgt 10 Prozentpunkte (Tabelle 6).

Tabelle 6: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:
Variation der Mengenbezeichnung "ein Fünftel"

Frage: *"Kürzlich war zu lesen, dass bei uns in Deutschland jede fünfte Frau raucht. Finden Sie, das ist viel, oder ist das nicht so viel?"*

In Parallelbefragungen hieß es: *"ein Fünftel der Frauen" / "20 Prozent der Frauen" / "eine von fünf Frauen"*.

	Anteil der Frauen vorgestellt als -			
	jede fünfte ("linear") %	20 Prozent ("prozentual") %	ein Fünftel ("geometrisch") %	eine von fünf ("gruppiert") %
Finde ich viel	55	55	50	48
Nicht so viel	39	41	47	49
Unentschieden	6	4	3	3
	$\overline{100}$	$\overline{100}$	$\overline{100}$	$\overline{100}$
n =	161	145	170	161

IfP-MP SS 2001

Aufgrund der vorliegenden Befunde kann als Zwischenergebnis festgehalten werden, dass die Verwendung numerischer Äquivalente bei der Beschreibung quantifizierter Sachverhalte einen erheblichen Einfluss auf die Wahrnehmung dieser Sachverhalte ausübt. Bei identischen Größenordnungen führen verschiedene Mengenbezeichnungen zu verschiedenen Einschätzungen der mit den Mengenbezeichnungen verbundenen Sachverhalte, was sich letztlich auch in der Bewertung dieser Sachverhalte niederschlagen dürfte.

Die getroffene Feststellung gründet sich auf eine Frageserie, in der die Befragten aufgrund des Eindrucks, den ihnen eine vorgestellte Größenordnung vermittelte, ein Urteil abgaben. Damit stellt sich die Frage, ob die Ergebnisse nur für solche Problemstellungen Gültigkeit

besitzen, oder ob sie auch bei vollkommen anders gelagerten Problemstellungen stabil sind. Zur Überprüfung dieser Frage wurde eine logisch anders aufgebaute Testfrage entwickelt, in der die Befragten ihre Vorstellungen von den - zuvor getrennt untersuchten - Größenordnungen klassifizieren sollten. Die Testfrage lautete: *"Was glauben Sie, wie viele Deutsche haben schon einmal Steuern hinterzogen? Würden Sie sagen ..."* Als Antwortkategorien, die ebenfalls vorgelesen wurden, standen die vier Größenordnungen "ein Fünftel, ein Viertel, ein Drittel oder die Hälfte der Deutschen" zur Auswahl. In den Parallelbefragungen wurden entsprechend die Mengenbezeichnungen der Größenordnungen variiert: "jeder fünfte, jeder vierte, jeder dritte oder jeder zweite Deutsche" usw.

Tabelle 7: Der Einfluss numerischer Äquivalente auf das Antwortverhalten:

Variation von Größenordnungen und Mengenbezeichnungen in einer einzigen Testfrage

Frage: *"Was glauben Sie, wie viele Deutsche haben schon einmal Steuern hinterzogen? Würden Sie sagen ein Fünftel, ein Viertel, ein Drittel oder die Hälfte der Deutschen?"*

In Parallelbefragungen hieß es: *"jeder fünfte, jeder vierte, jeder dritte oder jeder zweite Deutsche" / "20 Prozent, 25 Prozent, 33 Prozent oder 50 Prozent der Deutschen" / "einer von fünf, einer von vier, einer von drei oder einer von zwei Deutschen".*

Größen- ordnung	Mengenbezeichnung			
	"linear" %	"geometrisch" %	"gruppiert" %	"prozentual" %
Teiler 5	15	10	16	10
Teiler 4	18	20	17	17
Teiler 3	27	20	37	28
Teiler 2	40	50	30	45
	$\overline{100}$	$\overline{100}$	$\overline{100}$	$\overline{100}$
n =	143	157	157	158

"linear" = jeder fünfte, jeder vierte, jeder dritte, jeder zweite;

"geometrisch" = ein Fünftel, ein Viertel, ein Drittel, die Hälfte;

"gruppiert" = einer von fünf, einer von vier, einer von drei, einer von zwei;

"prozentual" = 20 Prozent, 25 Prozent, 33 Prozent, 50 Prozent.

IfP-MP SS 2001

Das Ergebnis ist eindeutig: die Vermutungen der Befragten über den Anteil der Deutschen, die schon einmal Steuern hinterzogen haben, fallen je nach den verwendeten Mengenbeschreibungen ganz verschieden aus. Insbesondere bei den Größenordnungen mit dem Teiler 2 (einer von zwei, die Hälfte, jeder zweite, 50 Prozent) und dem Teiler 3 (einer von drei, ein Drittel, jeder dritte, 33 Prozent) unterscheiden sich die Ergebnisse in den einzelnen

Versuchsgruppen um bis zu 20 bzw. 17 Prozentpunkte (Tabelle 7). Dies deutet darauf hin, dass der Einfluss unterschiedlicher Mengenangaben auf die Wahrnehmung und Bewertung quantifizierter oder zu quantifizierender Sachverhalte nicht auf bestimmte Problemstellungen beschränkt bleibt, sondern allgemeine Bedeutung besitzt.

4. Diskussion

Zahlen sind nicht kognitiv abstrakt, sondern besitzen je nach Darstellungsform verschiedene mentale Repräsentationen. Die mentalen Repräsentationen enthalten oder verzweigen auf semantische (arithmetische) und andere Informationen (z.B. Nachfolger von xy, Vorgänger von xy, gerade, ungerade usw.), mit denen kognitive Prozesse arbeiten (*Vorberg, Blankenberger* 1993). Dies trifft offenbar auch auf numerische Äquivalente zu, deren mentale Repräsentationen vermutlich unterschiedliche Mengenmuster enthalten (vgl. Schaubild). Möglicherweise sind diese Mengenmuster der Grund dafür, dass verschiedene Mengenbezeichnungen – unabhängig von Fragemodell und -formulierung, Frageinhalt, statistischer Größenordnung und Problemstellung – zu verschiedenen Einschätzungen der mit den Mengenbezeichnungen verbundenen Sachverhalte führen. Über alle Testfragen hinweg differenzieren die Ergebnisse in den verschiedenen Versuchsgruppen zwischen 10 und 21 Prozentpunkten.

Allerdings erfolgen die Einschätzungen der Befragten nicht einheitlich in der Weise, dass eine bestimmte Äquivalentform immer auch eine bestimmte Position innerhalb der – auf den jeweiligen Fragetext bezogenen – Rangfolge der Mengenbezeichnungen einnimmt. Beispielsweise ist mit der "linearen" Mengenbezeichnung (jeder Zweite, jeder Dritte, jeder Vierte usw.) nicht durchgehend der Eindruck eines größeren quantitativen Umfangs verbunden als mit der "gruppierten" Mengenbezeichnung (einer von zwei, einer von drei, einer von vier usw.). Dies ist offenbar eine Funktion der mit den mentalen Repräsentationen der numerischen Äquivalente verknüpften Informationen, über deren spezifische Einflussnahmen hier keine Aussagen getroffen werden können. Dennoch scheinen nach den vorliegenden Ergebnissen die numerischen Äquivalente nicht gleichwertig zu sein. Aus der Häufigkeit der Antworten, die anzeigen, dass ein gegebener Anteil als groß bzw. bedrohlich empfunden wird⁷, geht hervor, dass Mengenbezeichnungen, die im Vorstellungsraum als "lineare" und "geometrische" Mengenmuster abgerufen werden, die Quantität von Größenordnungen eher zu unterstreichen scheinen als Mengenbezeichnungen, die im Vorstellungsraum als "gruppierte" und "prozentuale" Mengenmuster abgerufen werden (Tabelle 8). Dies könnte bedeuten, dass sich mit bestimmten Mengenbezeichnungen Sachverhalte stärker oder schwächer konturieren lassen, was sich letztlich auch in der Bewertung dieser Sachverhalte niederschlägt.

7 Bezogen auf die Gesamtheit der logisch gleich aufgebauten Testfragen (Tabellen 2 bis 6).

Tabelle 8: Gibt es eine "Hierarchie" der Mengenbezeichnungen?

Häufigkeit der Antworten, die anzeigen, dass ein gegebener Anteil
als groß bzw. bedrohlich empfunden wird (Tabellen 2 bis 6)

		Mengenbezeichnung			
		"linear"	"geometrisch"	"gruppiert"	"prozentual"
		%	%	%	%
Es empfinden als bedrohlich, dass [...] der Arbeitnehmer sich schon mal krank meldet, ohne wirklich krank zu sein					
Teiler 2	(Tabelle 4)	53	58	37	40
Teiler 3	(Tabelle 2)	56	54	50	45
Teiler 3	(Tabelle 3)	52	63	60	53
Es empfinden als viel, dass [...] der Deutschen zu wenig Selbstvertrauen besitzt					
Teiler 4	(Tabelle 5)	49	45	38	34
Es empfinden als viel, dass [...] der Frauen in Deutsch- land raucht					
Teiler 5	(Tabelle 6)	55	50	48	55
Gesamt		<u>265</u>	<u>270</u>	<u>233</u>	<u>227</u>
Durchschnitt		53	54	47	45

Die Konsequenzen einer konturierenden Wirkung numerischer Äquivalente liegen auf der Hand. Ähnlich wie im Fall logischer Äquivalente könnte durch die – beabsichtigte oder unbeabsichtigte – Verwendung bestimmter Mengenbezeichnungen bei der Beschreibung statistischer Sachverhalte deren Bedeutung in den Augen der Adressaten herauf- oder herunterspielt werden. Eine solche Verzerrung, die lediglich auf der Art der Mengenbeschreibung beruht, ist im Bereich der Umfrageforschung doppelt relevant. Sie betrifft einerseits die Befragten, die auf den Fragestimulus entsprechend der verwendeten Mengenbezeichnung reagieren, und sie betrifft andererseits die Adressaten der Berichterstattung der Umfrageergebnisse, die durch die Verwendung numerischer Äquivalente bei der Darstellung der Ergebnisverteilung bestimmte Vorstellungen über die angesprochenen Größenverhältnisse entwickeln. Während die Bedeutung logischer Äquivalente in der Umfrageforschung seit den Methodenexperimenten von **Rugg** (1941) bekannt ist, blieb dieser Einfluss der numerischen Äquivalente bis heute verborgen.

5. Zusammenfassung

Ausgangspunkt der Untersuchung bildet die in der Umfrageforschung bekannte Tatsache, dass logisch äquivalente Begriffsverwendungen in der Frageformulierung einen erheblichen Einfluss auf die Wahrnehmung und Bewertung identischer Sachverhalte ausüben können. Ergebnisverzerrungen, die von der Verwendung numerischer Äquivalente in der Frageformulierung ausgehen, wurde hingegen trotz der gleich gelagerten Problematik bisher keine Aufmerksamkeit geschenkt. Im Sinne des von *Roper* (1940) vorgeschlagenen Forschungsprogramms zur systematischen Identifizierung von Schlüsselbegriffen, die im Fragebogen die subjektive Wahrnehmung und Bewertung des Frageinhalts durch die Befragten beeinflussen, wurde dieser Problemstellung anhand verschiedener Mengenbezeichnungen nachgegangen.

Mit Hilfe der Methode der gegabelten Befragung wurden für die Größenordnungen mit den Teilern 2, 3, 4 und 5 die entsprechenden Mengenangaben vierfach variiert. Im Falle der Größenordnung mit dem Teiler 2 lauteten die Mengenbezeichnungen "die Hälfte", "einer von zwei", "50 Prozent" sowie "jeder zweite", im Falle der Größenordnung mit dem Teiler 3 entsprechend "ein Drittel", "einer von drei", "33 Prozent" sowie "jeder dritte", usf. Ungeachtet der jeweiligen Größenordnung weichen die Ergebnisse je nach verwendeter Mengenbezeichnung deutlich voneinander ab. Die in mehreren Befragungen gefundenen Einzelergebnisse für die verschiedenen Größenordnungen wurden mit einer weiteren Testfrage überprüft, in die alle zuvor getrennt untersuchten Größenordnungen eingeschlossen waren. Der hochsignifikante Sammelbefund unterstreicht die anhand der Einzelbeobachtungen getroffenen Feststellungen. Unabhängig von der Größenordnung führt die Verwendung verschiedener Mengenbezeichnungen zu verschiedenen Einschätzungen der mit den Mengenangaben verbundenen Sachverhalte.

Die Einschätzungen erfolgen nicht einheitlich in dem Sinne, dass eine bestimmte Äquivalentform immer auch eine bestimmte Position innerhalb der – auf den jeweiligen Fragetext bezogenen – Rangfolge der Mengenbezeichnungen einnimmt. Über die getesteten Größenordnungen hinweg zeigt sich jedoch ein Übergewicht der "linearen" und "geometrischen" Mengenbezeichnungen auf den beiden vorderen Rangplätzen, wohingegen die "gruppierten" und "prozentualen" Mengenbezeichnungen überwiegend die beiden hinteren Rangplätze belegen. Vorbehaltlich weiterer Überprüfungen könnte dies darauf hindeuten, dass sich mit bestimmten Mengenbezeichnungen Sachverhalte stärker oder schwächer konturieren lassen, was sich über die unterschiedliche Wahrnehmung dieser Sachverhalte letztlich auch in ihrer Bewertung niederschlägt.

An einem der Medienberichterstattung entnommenen Beispiel lässt sich die Relevanz und zugleich die Problematik der Verwendung verschiedener Mengenbezeichnungen gut illustrieren. Bei der Darstellung der Situation der katholischen Kirche in Deutschland

(*FAZ* vom 24.11.1998) bedient sich der Verfasser im Bemühen um sprachliche Abwechslung aller Varianten von Mengenbezeichnungen. In dem Bericht heißt es: "Der Anteil der Katholiken an der Wohnbevölkerung in Deutschland beträgt nach der neuesten Statistik etwas mehr als ein Drittel (33,4 Prozent). In den alten Bundesländern ist zur Zeit noch etwa jeder vierte Mitglied der katholischen Kirche (40,7 Prozent) [sic!], in den neuen Ländern einschließlich Berlins einer von siebzehn (6,2 Prozent)." Die offenkundigen Schwierigkeiten bei der Verwendung der numerischen Äquivalente müssen nicht betont werden. Auch soll an dieser Stelle nicht darüber spekuliert werden, ob die "geometrische" (ein Drittel) und die "lineare" (jeder vierte) Mengenbezeichnung die tatsächliche Größenordnung stärker erscheinen lassen, während die "gruppierte" Mengenbezeichnung (einer von siebzehn) sie eher schwächer erscheinen lässt, zumal die Prozentwerte in Klammern stehen.

Das Beispiel wirft indessen eine Frage auf, die der weiteren Überprüfung bedarf. Treten die für die Größenordnungen mit den Teilern 2 bis 5 aufgezeigten Effekte auch bei anderen Größenordnungen auf, oder bleiben sie auf kleinere Teiler beschränkt? Sowohl die Einzelfragen zu den getesteten Größenordnungen als auch die Sammelfrage mit allen vier eingeschlossenen Größenordnungen deuten darauf hin, dass die Unterschiede zwischen den Mengenbezeichnungen bei kleineren Teilern größer ausfallen als bei größeren Teilern. Gibt es mithin eine "Grenze", ab der sich der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte nicht mehr oder aber anders auswirkt? Die "gruppierte" Mengenbezeichnung "einer von siebzehn" ist sicherlich nicht gerade anschaulich.

Die vorgestellten Ergebnisse entstammen der "Suchphase" in der Umfrageforschung, in der das Entdeckungspotential der Methode systematisch genutzt und Fährten aufgedeckt werden (*Noelle-Neumann, Petersen* 1996). Der Einfluss numerischer Äquivalente auf die Wahrnehmung und Bewertung identischer Sachverhalte ist sicherlich eine solche Fährte. Sie weist den Weg zu einem ganzen Forschungsprogramm, da die hiermit verknüpften Fragestellungen zu zahlreich sind, um in einem einzigen Wurf beantwortet zu werden.

Literatur

Besner, D., M. Coltheart 1979: Ideographic and alphabetic processing in skilled reading of English. In: *Neuropsychologia* 17, S. 467-472.

Bradburn, Norman M., Seymour Sudman 1989: *Polls and Surveys. Understanding What They Tell Us*. San Francisco, London.

Campbell, J.I.D., J.M. Clark 1988: An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman. In: *Journal of Experimental Psychology: General* 117, S. 204-214.

Campbell, J.I.D., J.M. Clark 1992: Cognitive number processing: An encoding complex perspective. In: *J.I.D. Campbell* (Hrsg.): *The nature and origin of mathematical skills*. Amsterdam, S. 457-492.

- Clark, J.M., J.I.D. Campbell** 1991: Integrated versus modular theories of number skills and acalculia. In: *Brain and Cognition* 17, S. 204-239.
- Clark, H.H., M.F. Schober** 1992: Asking questions and influencing answers. In: **J.M. Tanur** (Hrsg.): *Questions about questions. Inquiries into the cognitive bases of surveys*. New York, S. 15-48.
- Dehaene, Stanislas** 1999: *Der Zahlensinn oder Warum wir rechnen können*. Basel, Boston, Berlin.
- Dehaene, S., L. Cohen** 1991: Two mental calculation systems: A case study of severe acalculia with preserved approximation. In: *Neuropsychologia* 29, S. 1045-1074.
- Frankfurter Allgemeine Zeitung** (Nr. 273, 24.11.1998, S.6): [D.D.] Weniger Austritte, mehr Rückkehrer. Die katholische Kirche stellt einen neuen Trend fest.
- Friedrichs, Jürgen** 1973: *Methoden empirischer Sozialforschung*. Reinbek bei Hamburg.
- Gonzales, E.G., P.A. Kolers** 1982: Mental manipulation of arithmetic symbols. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, S. 308-319.
- Gonzales, E.G., P.A. Kolers** 1987: Notational constraints on mental operations. In: **G. Deloche, X. Seron** (Hrsg.): *Mathematical disabilities: A cognitive neuropsychological perspective*. Hillsdale, S. 27-42.
- Hinrichs, J.V., D.S. Yurko, Jing-Mei Hu** 1981: Two-digit number comparison: Use of place information. In: *Journal of Experimental Psychology: Human Perception and Performance* 7, S. 890-901.
- Hinrichs, J.V., J.L. Berie, M.K. Mosell** 1982: Place information in multidigit number comparison. In: *Memory and Cognition* 10, S. 487-495.
- Hörmann, Hans** 1983: The Calculating Listener, or How Many are *einige, mehrere* and *ein paar* (Some, Several, and a Few)? In: **Rainer Bäuerle, Christoph Schwarze, Arnim von Stechow** (Hrsg.): *Meaning, Use, and Interpretation of Language*. Berlin, New York, S. 221-234.
- Kahnemann, Daniel, Amos Tversky** 1984: Choices, Values, and Frames. In: *American Psychologist* 39, S. 341-350.
- Lienert, G.A., U. Raatz** 1994: *Testaufbau und Testanalyse*. Weinheim.
- Noelle-Neumann, Elisabeth, Thomas Petersen** 1996: *Alle, nicht jeder. Einführung in die Methoden der Demoskopie*. München.
- Payne, Stanley L.** 1965: *The Art of Asking Questions*. Princeton, New Jersey.
- Reuband, Karl-Heinz** 2001: "Erlauben" vs. "nicht erlauben" oder "verbieten"? Wie sich unterschiedliche Frage-Alternativen auf das Antwortverhalten auswirken. In: *ZA-Information* 48, S. 42-55.
- Roper, Elmo** 1940: Wording Questions for the Polls. In: *Public Opinion Quarterly* 4, S. 129-130.
- Rugg, D.** 1941: Experiments in wording questions: II. In: *Public Opinion Quarterly* 5, S. 91-92.
- Schumann, H., S. Presser** 1981: *Questions and answers in attitude surveys. Experiments in question form, wording and context*. New York.
- Sudman, Seymour, Norman M. Bradburn** 1982: *Asking Questions. A Practical Guide to Questionnaire Design*. San Francisco, Washington, London.
- Tennstädt, Friedrich, Jochen Hansen** 1982: Validating the recency and through-the-book techniques. In: **Harry Henry** (Hrsg.): *Readership Research: Theory and Practice. Proceedings of the first International Symposium*, New Orleans, London, S. 106-121.
- Vorberg, Dirk, Sven Blankenberger** 1993: Mentale Repräsentationen von Zahlen. In: *Sprache und Kognition* 12, S. 98-114.